

Experiencias en la edición digital de textos con codificación XML. Mejoras en las bases de datos del *Arquivo Galicia Medieval*

MIGUEL ÁNGEL POUSADA CRUZ¹

*Centro Ramón Piñeiro para a Investigación en Humanidades
Universidade de Santiago de Compostela*

Resumen: El objetivo de este trabajo es presentar las últimas mejoras que han sido realizadas en la interfaz gráfica de usuario de la base de datos *MedDB* (*Arquivo Galicia Medieval* – CRPIH).

Palabras clave: Filología románica, Humanidades digitales, bases de datos.

Abstract: The aim of this paper is to review the latest improvements that have been made in the GUI of the *MedDB* database (*Arquivo Galicia Medieval* – CRPIH).

Key words: Romance Philology, Digital Humanities, databases.

1. INTRODUCCIÓN

EL CENTRO RAMÓN PIÑEIRO para a Investigación en Humanidades es un centro de investigación auspiciado por la *Secretaría Xeral de Política Lingüística* de la *Xunta de Galicia* que se encarga de desarrollar proyectos de investigación sobre estudios literarios, lingüísticos, históricos y antropológicos vinculados con Galicia. Dentro de las líneas de investigación relacionadas con el ámbito de la Informática humanística y de la Filología románica, se encuentra el macroproyecto *Arquivo Galicia Medieval* (ARGAMED), dirigido por la Prof.^a Mercedes Brea, coordinadora del grupo de investigación *GI-1350 Románicas (Filoloxía e literatura medieval)* de la Universidad de Santiago de Compostela.

El objetivo fundamental de este macroproyecto es poner a disposición del público interesado no solo los textos literarios medievales escritos en romance peninsular occidental, sino también los resultados derivados de ellos, a través, entre otras vías, de diversas bases de datos, consultables en línea, que incorporan la información mediante archivos que emplean el lenguaje de marcas XML².

Como ya hemos comentado en otras ocasiones el archivo trabaja en la actualidad en el desarrollo y mejora de tres bases de datos (*MedDB*, *BiRMED* y *PalMED*), de acceso gratuito. La información aportada en conjunto por estas tres bases (respectivamente, filológica, bibliográfica y paleográfica) ayuda al usuario a entender las obras incluidas en esta tradición literaria romance desde una perspectiva poliédrica, al circunscribirse estas herramientas informáticas (y los trabajos del macroproyecto) dentro de los ámbitos de la Filología románica y de las Humanidades digitales, esto es, de

¹ Trabajo de investigación realizado por el autor con el apoyo de una beca de formación del *Centro Ramón Piñeiro para a Investigación en Humanidades*, financiado por la *Secretaría Xeral de Política Lingüística* de la *Xunta de Galicia* y que se engloba en el marco de los proyectos de investigación *Lírica Profana Galego-Portuguesa*, perteneciente al macroproyecto *Arquivo Galicia Medieval* (CRPIH), y *La expresión de las emociones en la lírica románica medieval* (FFI2012-37355), financiado por el Ministerio de Economía y Competitividad. Ambos proyectos están dirigidos por la Dr.^a Mercedes Brea, Catedrática de Filología Románica de la USC.

² El macroproyecto está formado por tres proyectos concretos (*Cantigas de Santa Maria*, *Lírica Profana Galego-Portuguesa* y *Prosa Literaria Medieval Galega*). Para una breve descripción, *vid.* Pousada Cruz (2011). Más información sobre las bases en Domínguez Carregal, López Macías y Vega Vázquez (2009) y Domínguez Carregal y Vega Vázquez (2009).

disciplinas científicas que son interdisciplinares y transversales, desde el punto de vista de su metodología. Además, la vinculación del archivo con otros proyectos similares del ámbito comunitario³ permite analizar y comprender la complejidad de cada uno de estos textos en relación con el contexto lingüístico y literario románico europeo.

2. MEDDB-BASE DE DATOS DA LÍRICA PROFANA GALEGO-PORTUGUESA

Las bases de datos juegan un papel muy importante en el seno del ARGAMED. La unión de filólogos e informáticos permitió desarrollar bases de datos que fueron pioneras dentro del ámbito peninsular y sobre las cuales todavía se sigue trabajando.

El hecho de entender el mantenimiento de estos bancos textuales como una prioridad ha permitido que ARGAMED continúe siendo un exponente clave dentro de los estudios sobre la lírica medieval gallego-portuguesa. Las continuas adaptaciones y mejoras en la interfaz y la revisión periódica de la información contenida en las bases ha permitido cumplir el objetivo del macroproyecto, incrementándose en los últimos años, además, el número de usuarios registrados y de consultas.

Como es ya bien conocido *MedDB* es una base de datos que contiene el corpus completo de las cantigas medievales de los trovadores gallego-portugueses (1691 textos acompañados de información filológica y literaria).

En el año 1996, el equipo investigador del proyecto publica *Lírica Profana Galego-Portuguesa (LPGP)*. Por primera vez, el corpus lírico trovadoresco profano gallego-portugués podía ser consultado en una única obra, sin necesidad de acudir a distintas fuentes. Para completar el corpus fueron seleccionados los 1691 textos de entre las ediciones y estudios realizados hasta aquella fecha, atendiendo a criterios de rigor filológico. La presentación crítica de cada texto fue acompañada de información retórico-formal y bibliográfica, y se realizó una aproximación a los datos biográficos conocidos hasta aquel momento sobre cada autor.

Dos años más tarde, en 1998, se publicó la primera versión de *MedDB*, tomando como base los textos presentes en *LPGP* y empleando el lenguaje de marcas SGML. El resultado fue un banco textual pionero en el uso de los lenguajes de marcas en el ámbito peninsular y que tuvo una vida útil de diez años.

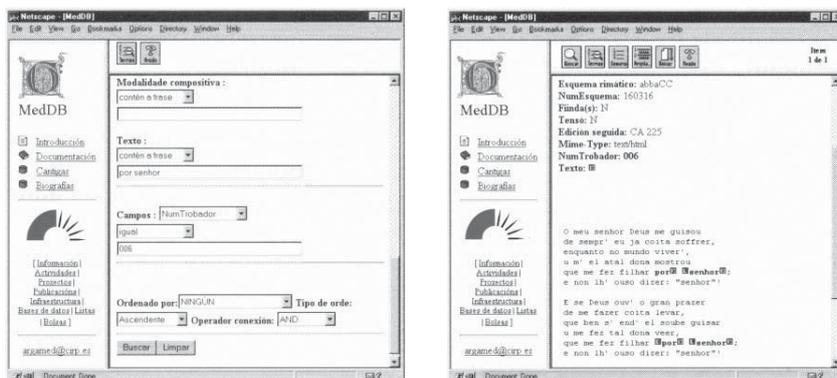


Imagen 1. Pantallas de búsqueda y de resultados de MedDB1

³ V. gr. el proyecto internacional *Lírica Europea*, coordinado desde la *Università di Roma-La Sapienza* por el Prof. Roberto Antonelli y cuyo objetivo es reunir bajo una única base de datos los diversos *corpora* de la lírica medieval o el proyecto *Lessico europeo dell'affettività* que prevé la realización de un nuevo modelo de base de datos semántica —vinculada con la esfera de los afectos— que permitirá realizar búsquedas por las categorías emocionales (*Laetitia, Tristitia, Timor, Ira, Cupiditas*) que contienen las distintas emociones a las que están asociados los lemas de cada una de las tradiciones líricas. El *Centro Ramón Piñeiro para a Investigación en Humanidades* es miembro, además, de *Medioevo Europeo - COST Action IS1005 (Medieval Cultures and Technological Resources)*, una acción apoyada por *COST (European Cooperation in Science and Technology)*, dentro del *7 Programa Marco* de la UE, y gestionada por la *ESF (European Science Foundation)*.

Con esta base de datos en línea, el equipo de investigadores —coordinado por el personal informático del centro—, continuó desarrollando la segunda versión de la base (la actual MedDB2): la información fue codificada empleando, esta vez, el lenguaje de marcas XML y se renovó el sistema de gestión de la base de datos. El resultado fue una nueva herramienta informática, con una interfaz de usuario más amigable e intuitiva, que ampliaba, además, las posibilidades de interrogación (cuatro grandes vías de acceso —por cantiga, por estrofa, por verso y por trovador— con sus correspondientes y diversos campos y filtros) y de recuperación de la información almacenada. Además de incorporar las imágenes digitalizadas de los testimonios, fue revisado el contenido de la base, ya que habían aparecido nuevas ediciones —algunas realizadas por miembros del equipo de ARGAMED y del grupo de investigación *GI-1350 Románicas (Filoloxía, Literatura medieval)* de la USC— y se habían aclarado algunos problemas de atribución de determinadas cantigas a distintos autores. En el año 2008, sale a la luz MedDB2.



Imagen 2. Pantalla de inicio de MedDB2

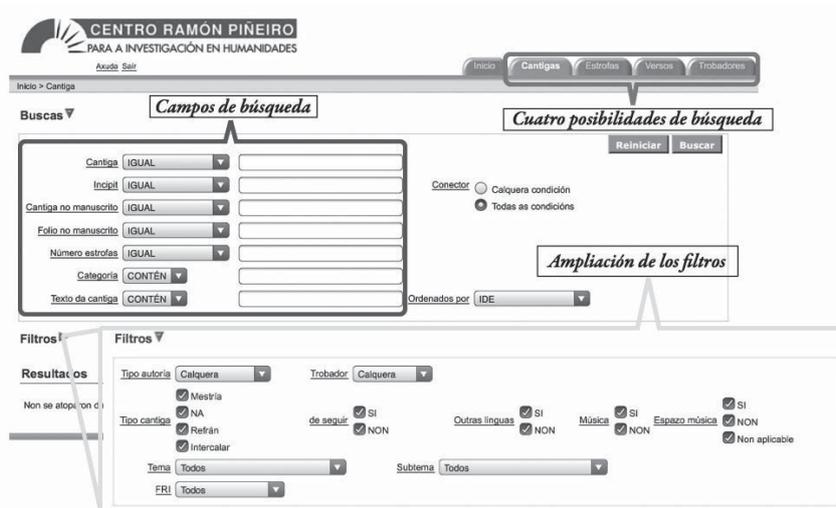


Imagen 3. Detalle de las posibilidades de búsqueda, campos y filtros de MedDB2, desde la búsqueda por cantiga

Poco a poco, aquella primera base de datos textual *MedDB1* (1998), realizada sobre *LPGP* (1996), dio paso a un nuevo banco de datos, *MedDB2* (2008), que cada vez se parecía más a una edición digital que permitía tener en una misma pantalla presentación crítica del texto e imagen de los manuscritos que lo transmiten. La pronta publicación de *PalMED* permitirá, además, disponer de la transcripción paleográfica de todos los testimonios conservados.

Imagen 4. Pantalla de resultados de *MedDB2* desde «búsqueda por cantiga»

En la actualidad, el equipo investigador del macroproyecto trabaja en la tercera versión de la base (*MedDB3*), que incorporará nuevos textos editados por el grupo de investigación *GI-1350 Románicas (Filoloxía, Literatura Medieval)* a lo largo de estos años, así como nuevas mejoras en la interfaz de usuario, haciéndola, como podemos ver en las siguientes capturas, más amigable: se ha modificado la posición de los filtros, se han posibilitado nuevos campos de búsqueda, etc.

Imagen 5. Pantalla provisional inicial de *MedDB3*

Además de un pequeño cambio en la imagen inicial, el usuario puede elegir, en la parte superior derecha, entre *búsqueda simple* —dirigida a usuarios no especialistas, de acceso libre y que permite obtener velozmente una composición concreta de un trovador (a través del listado de versos de *incipit*), desprovista de más información, o el conjunto completo de textos de un autor (a través del campo *autor*)— y *búsqueda avanzada* —dirigida a especialistas y/o interesados en obtener más información de un texto concreto, de un autor específico, etc., de acceso gratuito previo registro automático—; esta última *opción de búsqueda* permite interrogar el banco de datos desde las

cuatro vías ya presentes en MedDB2 —por *cantiga*, por *estrofa*, por *verso* e por *trovador*—. Además, la nueva versión permitirá buscar información de diversa índole que no estaba operativa en las versiones anteriores —rúbricas y apostillas marginales—.



Imagen 6. Pantalla provisional de resultados por «búsqueda simple (cantiga)»

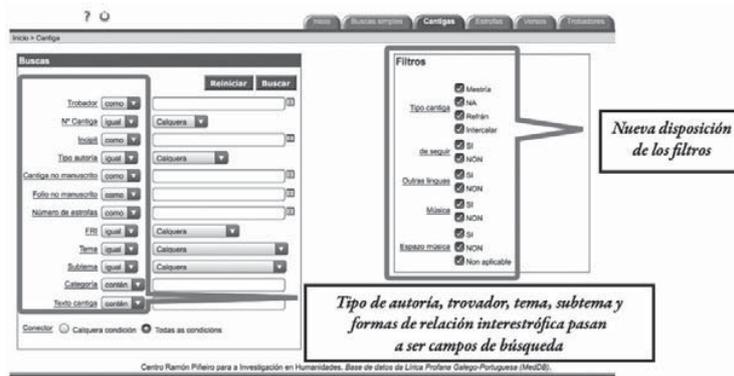


Imagen 7. Pantalla provisional de consulta a través de «búsquedas avanzadas (cantiga)»

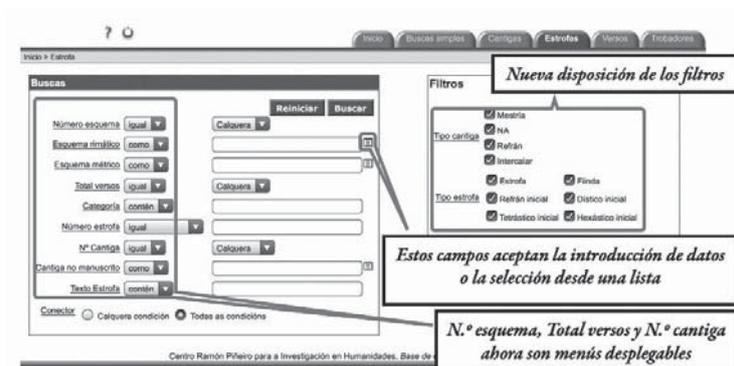


Imagen 8. Pantalla provisional de consulta a través de «búsquedas avanzadas (estrofa)»

Imagen 9. Pantalla provisional de consulta a través de «búsquedas avanzadas (verso)»

Imagen 10. Pantalla provisional de resultado específico a través de «búsquedas avanzadas (autor)»

3. CONCLUSIÓN

Como puede observarse en las capturas anteriores, las bases de datos del *Arquivo Galicia Medieval* y, en concreto, *MedDB-Base de datos da Lírica Profana Galego-Portuguesa*, son herramientas informáticas útiles para el estudio filológico de este corpus literario y están en continuo proceso de renovación y revisión. La evolución en la elección del lenguaje de marcado dentro de los procesos de codificación de los textos en las distintas versiones muestra, también, la preocupación de todo el equipo por el uso de estándares internacionales que permitan que los datos contenidos en la base puedan ser más *accesibles* y *perdurables*. La revisión constante de la información almacenada y el consecuente cambio de ediciones que han quedado obsoletas por nuevos textos editados bajo los parámetros de la Crítica Textual moderna ayudan a mantener el *rigor intelectual* exigible para un corpus filológico de estas características y de este nivel académico. *Acceability*, *longevity* y *intellectual integrity*, usando las palabras de Sperberg-MacQueen (1994), son, por tanto, los pilares sobre los que pretendemos asentar los proyectos que realizamos y desarrollamos dentro de *Arquivo Galicia Medieval*, con el objetivo final de mirar siempre hacia nuestros compañeros europeos, buscando no

solo la internacionalización de nuestra investigación, sino también el intercambio científico en un intento de situar los estudios sobre literatura medieval gallego-portuguesa en el lugar que les corresponde dentro del ámbito de la Filología Románica y de las Humanidades Digitales.

REFERENCIAS BIBLIOGRÁFICAS

Fuentes primarias (corpus y bases de datos)

- BREA, Mercedes (coord.) (1996): *Lírica Profana Galego-Portuguesa. Corpus completo das cantigas medievals, con estudo biográfico, análise retórica e bibliografía específica*. Santiago de Compostela: Centro Ramón Piñeiro para a Investigación en Humanidades-Xunta de Galicia.
- BiRMED2-BIBLIOGRAFÍA DE REFERENCIA DO ARQUIVO GALICIA MEDIEVAL. Versión 2.1. Santiago de Compostela: Centro Ramón Piñeiro para a Investigación en Humanidades [en línea], <<http://www.cirp.es/pls/bdo2/f?p=birmed2>> [Consultado en julio de 2013].
- MEDDB2-BASE DE DATOS DA LÍRICA PROFANA GALEGO-PORTUGUESA. Versión 2.3.3. Santiago de Compostela: Centro Ramón Piñeiro para a Investigación en Humanidades [en línea], <<http://www.cirp.es/pls/bdo2/f?p=meddb2>> [Consultado en julio de 2013].
- PALMED-BASE DE DATOS PALEOGRÁFICA DA LÍRICA PROFANA GALEGO-PORTUGUESA. Versión 1.0. Santiago de Compostela: Centro Ramón Piñeiro para a Investigación en Humanidades [en prensa].

Fuentes secundarias

- DOMÍNGUEZ CARREGAL, Antonio Augusto, Marta LOPEZ MACÍAS e Isabel VEGA VÁZQUEZ (2009): «BiRMED: Bibliografía de Referencia da Lírica Medieval Galego-Portuguesa». Furio Brugnolo y Francesca Gambino (eds.), *La lirica romanza del Medioevo. Storia, tradizioni, interpretazioni. Atti del VI convegno triennale della Società Italiana di Filologia Romanza (Padova-Stra, 27 settembre – 1 ottobre 2006)*. Padova: UNIPRESS.
- DOMÍNGUEZ CARREGAL, Antonio Augusto e Isabel VEGA VÁZQUEZ (2009): «MedDB2: Base de datos da Lírica Profana Galego-Portuguesa». J. Cañas Murillo y Francisco J. Grande Quejigo (eds.), *Medievalismo en Extremadura: Estudios sobre Literatura y Culturas Hispánicas de la Edad Media*. Extremadura: Servicio de publicaciones de la Universidad de Extremadura, 727-731.
- POUSADA CRUZ, Miguel Ángel (2011): «Filología y mundo digital. La explotación de los recursos digitales en la edición de los textos romances». María Angustias Beas Teruel (ed.), *Nuevas líneas de investigación en el estudio diacrónico de la lengua española*. Palma de Mallorca: Edicions Universitat de les Illes Balears, 281-292.
- SPERBERG-MCQUEEN, C. Michael (1994): «Textual Criticism and the Text Encoding Initiative». *Annual Convention of the Modern Language Association* [en línea], <<http://www.tei-c.org/Vault/XX/mla94.html>> [Consultado en julio de 2013].